Documentation for NSHAPC Public Use Data Files
Urban Institute, Washington, DC   August 1999

Explanation of Tests Used to Determine Statistical Significance

In the course of working on the summary and technical reports of the NSHAPC project, many instances arose where the significance of a certain percentage or number had to be tested. If the data in question did not meet the significance guidelines outlined by the Bureau of the Census, it was not discussed in either of the NSHAPC reports. This docuemnt gives the formulas used to calculate statistical significance for the NSHAPC reports. A standard of *alpha* = .10 and a design effect of 3 were used in all cases.

Case 1: A single percentage not being compared to another percentage. An example of this case is "54 percent of all NSHAPC clients were homeless at the time of the survey." The following formula was used to establish a confidence interval around this percentage.

$$=1.645*SQRT(ABS((P_1*(1-P_1))/N_1)*3)$$

Where P_1 = the percentage, expressed in decimal notation, N_1 = the N for the percentage, and 3 = the design effect. The resulting number represents the $\pm$ value of the confidence interval around this percentage.

Case 2: Two percentages are being compared to one another. An example of this case is "68 percent of currently homeless clients are men compared to 54 percent for formerly homeless clients." The formula to determine whether or not this comparison is statistically significant is:

$$=(ABS(P_1-P_2))/(SQRT((P_1*(1-P_1)*3)/N_1)+((P_2*(1-P_2)*3)/N_2))$$

Where P_1 and P_2 = the two percentages, expressed in decimal notation, N_1 and N_2 = the Ns for the two percentages, and 3 = the design effect. If the resulting number is greater than or equal to 1.645, the comparison meets the 90% criterion for statistical significance. If the resulting number is less than 1.645, the comparison is not significant at the 90% level.

Case 3: A number (estimate) is reported and it is necessary to establish a confidence interval around this number. An example of this would be "there were approximately 9,000 soup kitchens in the United States at the time of the NSHAPC survey." The following formula was used to calculate the confidence interval around a number.

$$=1.645*(Q_1(SQRT(3/N_1)))$$

Where Q_1= the number of programs/providers of a given type (soup kitchens, in the above example) and N_1= the total number of programs/providers in the whole sample. The design effect is still equal to three. The resulting number is the $\pm$ confidence interval associated with the estimated number of programs.